Geophysical Research Abstracts
Vol. 16, EGU2014-12909-2, 2014
EGU General Assembly 2014


# Ontology Based Vocabulary Matching for Oceanographic Instruments

Yu Chen (1), Adam Shepherd (2), Cyndy Chandler (2), Robert Arko (3), and Adam Leadbetter (4)

(1) Rensselaer Polytechnic Institute, Department of Computer Science, Tetherless World Constellation, Troy, United States (cheny18@rpi.edu, 5185227669), (2) Woods Hole Oceanographic Institution, Marine Chemistry, BCO-DMO, Woods Hole, Massachusetts, United States of America, (3) Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, United States of America, (4) British Oceanographic Data Centre, Liverpool, United Kingdom,

Data integration act as the preliminary entry point as we enter the era of big data in many scientific domains. However the reusefulness of various dataset has met the hurdle due to different initial of interests of different parties, therefore different vocabularies in describing similar or semantically related concepts. In this scenario it is vital to devise an automatic or semi-supervised algorithm to facilitate the convergence of different vocabularies.

The Ocean Data Interoperability Platform (ODIP) seeks to increase data sharing across scientific domains and international boundaries by providing a forum to harmonize diverse regional data systems. ODIP participants from the US include the Rolling Deck to Repository (R2R) program, whose mission is to capture, catalog, and describe the underway/environmental sensor data from US oceanographic research vessels and submit the data to public long-term archives. In an attempt to harmonize these regional data systems, especially vocabularies, R2R recognizes the value of the SeaDataNet vocabularies served by the NERC Vocabulary Server (NVS) hosted at the British Oceanographic Data Centre as a trusted, authoritative source for describing many oceanographic research concepts such as instrumentation.

In this work, we make use of the semantic relations in the vocabularies served by NVS to build a Bayesian network and take advantage of the idea of entropy in evaluating the correlation between different concepts and keywords. The performance of the model is evaluated against matching instruments from R2R against the SeaDataNet instrument vocabularies based on calculated confidence scores in the instrument pairings. These pairings with their scores can then be analyzed for assertion growing the interoperability of the R2R vocabulary through its links to the SeaDataNet entities.